

# Genomic Structure of Proton-Coupled Oligopeptide Transporter *hPEPT1* and pH-Sensing Regulatory Splice Variant

Submitted: October 23, 2000; Accepted: January 19, 2000; Published: February 6, 2000.

Arto Urtti ac

Department of Biopharmaceutical Sciences, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143-0446 and Department of Pharmaceutics, University of Kuopio, POB 1627, 70211 Kuopio, Finland

Wolfgang Sadée

Department of Biopharmaceutical Sciences, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143-0446

Susan Jean Johns

Sequence Analysis & Consulting Service, Computer Graphics Laboratory, Department of Pharmaceutical Chemistry, University of California San Francisco, 513 Parnassus Ave, San Francisco, CA 94143-0446.

**ABSTRACT** Proton-coupled oligopeptide transporter PEPT1 facilitates the transport of dipeptides and peptoid drugs (including antibiotics) across the cell membranes of endothelial and epithelial cells. Substrate transport by the proton symport is driven by pH gradients, while the profile of pH sensitivity is regulated by a closely related protein, hPEPT1-RF. We investigated the genomic structure of *hPEPT1* and *hPEPT1-RF*. Analysis of the high-throughput genomic sequence (HTGS) database revealed that hPEPT1 and hPEPT1-RF are splice variants encoded by the same gene located in chromosome 13, consisting of 24 exons. hPEPT1 is encoded by 23 exons and hPEPT1-RF by 6 exons. Coding sequences of hPEPT1-RF share 3 exons completely and 2 exons partially with hPEPT1. The genomic organization of *hPEPT1* shows high similarity with its mouse orthologue. Exon-intron boundaries occur mostly in the loops connecting transmembrane segments (TMSs), suggesting a modular gene structure reflecting the TMS-loop repeat units in hPEPT1. The putative promoter region of *hPEPT1* contains TATA boxes and GC-rich regions and a potential insulin responsive element.

**KeyWords:** Peptide transporter, Genomic structure, Splice variant

## INTRODUCTION

Proton-coupled oligopeptide transporters (POTs) comprise the transport family 2.A.17 (for transporter classification see <http://www.biology.ucsd.edu/~msaier/transport/titlepage.html>). Oligopeptide transporters are symporters driven by the flux of protons; they have a molecular architecture consisting of ~12 predicted TMSs (1). Members of the POT family include peptide transporter 1 (PEPT1) (2, 3), peptide transporter 2

(PEPT2) (4), peptide/histidine transporter 1 (PHT-1) (5, 6), and peptide/histidine transporter 2 (PHT-2) (6). Recently, a cDNA termed *PET3* (NM\_016582), which is largely identical to PHT-2, has been deposited into the nr database (<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi>).

The peptide transporter 1 gene of rabbits was cloned in 1994 (2), and the human orthologue (*hPEPT1*) was cloned shortly after (3). Human *PEPT1* cDNA contains 3105 base pairs (bp), and the predicted protein consists of 708 amino acids. The transporter protein has 12 predicted TMSs and 2 putative protein kinase C phosphorylation sites. The membrane topology of the human dipeptide transporter, hPEPT1, was determined by epitope insertions by Covitz et al (7). PEPT1 is expressed in the intestine (brush border), early proximal kidney tubuli, liver, placenta, and pancreas (3, 8). In the intestines, PEPT1 facilitates absorption of digested dipeptides so that most of the dietary nitrogen is absorbed as dipeptides rather than as amino acids (9).

Human PEPT1 has broad substrate specificity. The substrates include di- and tripeptides and peptoid drugs. Thus, PEPT1 mediates the high bioavailability of many hydrophilic beta-lactam antibiotics (10). In addition, PEPT1 is suggested to play a role in intracellular peptide transport, including lysosomal transport (11).

Saito et al (12) have described a highly related transcript, termed hPEPT1-RF, which modulates the activity of human PEPT1. The cDNA for the regulatory factor encodes an open reading frame of 208 amino acids. Residues 18-195 are identical to residues 8-185 in hPEPT1, while sequences 1-17 and 196-208 are unique. Both hPEPT1 and hPEPT1-RF

**Corresponding author:** Arto Urtti, Ph.D.; Department of Pharmaceutics, University of Kuopio; POB 1627, 70211 Kuopio, Finland. Telephone: +358 17 162480; Fascimile: +358 17 162252; E-mail: [Arto.Urtti@messi.uku.fi](mailto:Arto.Urtti@messi.uku.fi)

are expressed in Caco-2 cells. Expression studies in *Xenopus* oocytes and Caco-2 cells showed that the regulatory factor shifted the pH-sensitivity profile of hPEPT1-mediated peptide transport (12).

Although somatic cell hybrid analysis and in situ hybridization studies of Liang et al (3) positioned hPEPT1 to chromosome 13 q33-q34, the genomic structures of human PEPT1 and hPEPT1-RF were not known. Genomic organization of the mouse PEPT1 gene has been reported recently (13) as having a length of 38 kb with 23 exons.

The aim of our study was to determine the genomic structure of hPEPT1 and hPEPT1-RF. We identified a common gene with 24 exons encoding both hPEPT1 and the regulatory factor in clones representing chromosome 13. hPEPT1 and hPEPT1-RF are splice variants of the same gene.

## MATERIALS AND METHODS

Advanced (BLAST) analysis was carried out using the National Center for Biotechnology Information (NCBI) Web server (<http://ncbi.nlm.nih.gov>). BLOSUM62 matrix was used with default parameters. The analysis was done with and without filtering of the low-complexity sequences and without masking of repetitive elements. Queries used the cDNA sequences of human PEPT1 (accession number: NM\_005073) and hPEPT1-RF (AB001328) and the high-throughput genomic sequence (HTGS) database.

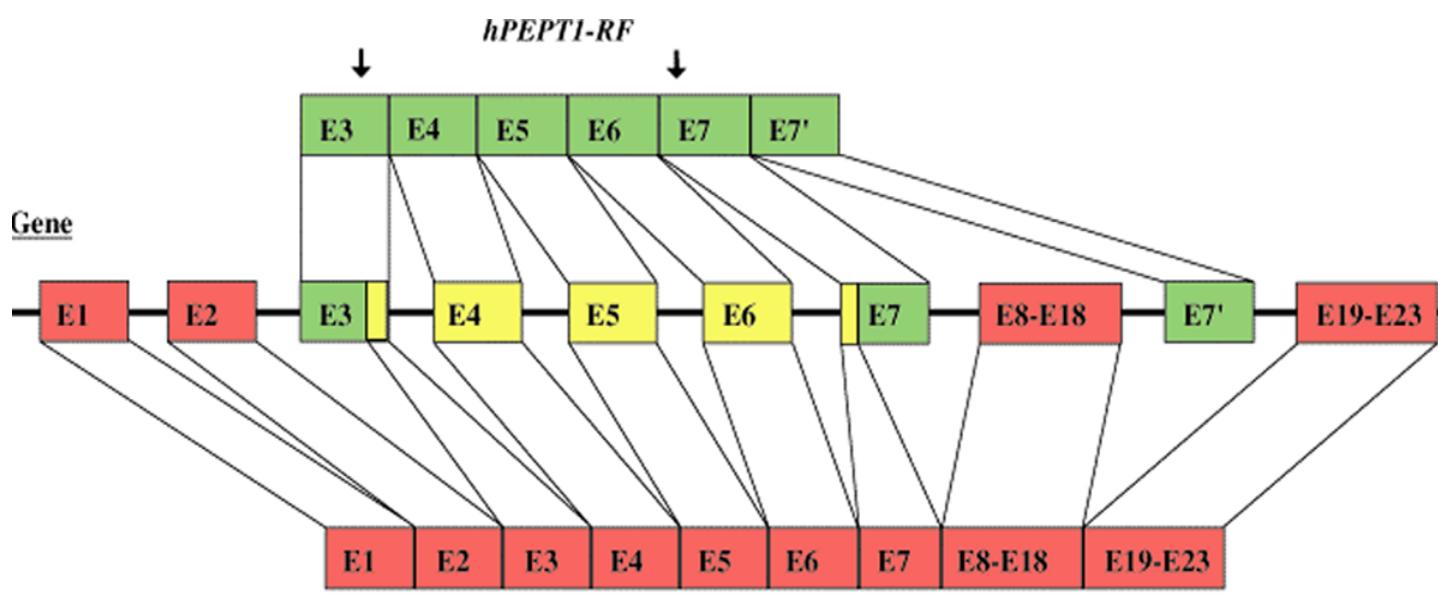
Using the accession number of the mRNA sequence, we retrieved the (CDS) sequence from NCBI and performed a BLAST search of the HTGS database.

Results were filtered using the blastfl.py code written by Arne Mueller (BLAST2 Parser ver. 1.2, ©Arne Mueller, <http://www.bmm.icnet.uk/people/mueller>. The obtained hits were filtered to ensure that only data from the same species and chromosome were used. The alignments served to locate the exons in the genomic sequence. When problems arose, the sequences were examined by hand to attempt to resolve or identify possible alternative splice sites. Membrane topological prediction was done using the (TOPPRED) program.

The sequences 2 kb upstream from the transcription start sites of hPEPT1 and hPEPT1-RF were investigated using programs FindPatterns and FitConsensus (Genetics Computer Group, Madison, WI) to locate possible promoters and enhancer sites.

## RESULTS

Bioinformatic analysis revealed that hPEPT1 and hPEPT1-RF are encoded by the same gene located in chromosome 13, clone RP11-56D6 (accession: AL357553). hPEPT1 contains 23 exons (Table 1, Figure 1), and hPEPT1-RF contains 6 exons (Table 2, Figure 1). Human PEPT1 and hPEPT1-RF share 3 exons completely, and 2 exons are partially shared (Figure 1). Therefore, hPEPT1-RF and hPEPT1 are splice variants of the same gene that has in total 24 exons. Over the course of the study, additional genomic clones became available containing all hPEPT1 exons in several contiguous fragments, and these served to verify the order and intronic sizes provided in Tables 1 and 2.



**Figure 1.** Gene structure and splicing of the hPEPT1 gene. hPEPT1 and hPEPT1-RF completely share exons 4-6 and partially share exons 3 and 7, where the alternative splice sites are located (indicated by the vertical lines). Exon 7' represents a repetitive element. Arrows indicate the translation start and stop sites.

hPEPT1

To view the exonic sequences, see appendix.

Table 1. hPEPT1 gene.

EXON		INTRON			EXON			
No.	Size (bp)	3' junction	5' junction	No.	3'junction	5' junction	No.	Phase
1	62	gtacgcctcg...	gttccatgg...	1	...tttttttag	GAATGTCCA..	2	1
2	17	gtgaggattacc...	gttccatgg...	2	...ttttgtcag	AGTTTCTTT..	3	0
3	200/82 <sup>b</sup>	gttaactgtta...	gttccatgg...	3	...ctgcaaaaag	CAATCCTGTA..	4	1
4	142	gttgagggtggc...	gttccatgg...	4	...gaaacacag	GACCATTGT..	5	2
5	120	gttgagggtgg...	gttccatgg...	5	...ctttccatgg	GGTGCTGTGTC..	6	2
6	100	gttcaagggtta...	gttccatgg...	6	...tcattgtcag	GAGAACACAA..	7	0
7	132/91 <sup>b</sup>	gttcaaggat...	gttccatgg...	7	...tctccccatgg	TTCAAACAAAT..	8	1
8	84	gttcaaggatgg...	gttccatgg...	8	...cctatccatgg	TTGTGTTTG..	9	1
9	83	gttcaaggatgtt...	gttccatgg...	9	...ttccatccatgg	TTTGGCCATC..	10	0
10	87	gttcaaggtaac...	gttccatgg...	10	...caactacatgg	GAGGGGGCTC..	11	0
11	88	gttcaaggtaaa...	gttccatgg...	11	...atttttaatgg	GGCTCCAGG..	12	0
12	45	gttcaaggatgtt...	gttccatgg...	12	...tttttttag	GGAGGCTCTT..	13	0
13	33	gttcaaggatgtt...	gttccatgg...	13	...cctctgtcag	ACCGTGAAC..	14	0
14	89	gttcaaggatgtt...	gttccatgg...	14	...tgcctatgg	CTCCCTTGAA..	15	2
15	82	gttcaaggatgtt...	gttccatgg...	15	...tttgctcag	AAAACCTCTT..	16	0
16	120	gttcaaggatgtt...	gttccatgg...	16	...cctctatgg	ACAAATGCA..	17	0
17	147	gttcaaggatgtt...	gttccatgg...	17	...ctctctcag	GTAAAGGAT..	18	0
18	50	gttcaaggatgtt...	gttccatgg...	18	...ctgttttcag	ATTGGTAAA..	19	2
19	108	gttcaaggatgtt...	gttccatgg...	19	...ttttgtatgg	AAAAAGGCTT..	20	2
20	109	gttcaaggatgtt...	gttccatgg...	20	...actgtcag	AATGACAGC..	21	0
21	144	gttcaaggatgtt...	gttccatgg...	21	...cttcctcag	GCTCCCTTCC..	22	0
22	108	gttcaaggatgtt...	gttccatgg...	22	...tcttccaatgg	TGGGCCGGAG..	23	0
23	1111							

Exon boundaries are shown in upper case and intron boundaries in lower case. Codon phase refers to the codon in the 5' end of the exon. Uncertain sizes of introns are indicated by > sign due to the boundary between two fragments of the clone in HTGS database. <sup>b</sup>Size of whole exon/ size of the part spliced into hPEPT1 mRNA.

## MEMBRANE TOPOLOGICAL PREDICTION OF hPEPT-1

## MEMBRANE TOPOLOGICAL PREDICTION OF hPEPT1-RF

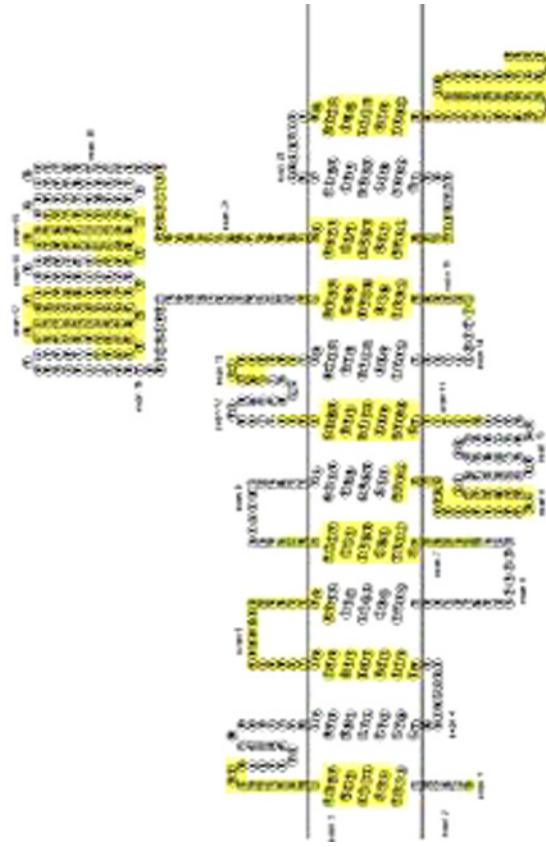


Figure 2. Membrane topology prediction for hPEPT1. The prediction was carried out using the TOPPRED program. The exon boundaries are indicated by alternating color codes.

TABLE 2. Exon-intron organization of hPEPT1-RF gene.

No.	EXON		INTRON		EXON	
	Size (bp)	3' junction	5' junction	Size (bp)	3' junction	5' junction
3	200	...GAATGCGAG	gtactgt...	104	...ctgaaaaag	CAATCCTGA
4	142	...AAAGTTCAA	gtgatggc...	2092	...gaaacacag	GACCATTTGT
5	120	...TGTGCACGT	gtgatgtgg...	2009	...cttcccg	GGTGCTGTC
6	100	...GAGGCCAG	gtaggta...	219	...tcattgcag	GAGAACACA
7	1054	...TTTTTTAG		5246	7	3' UTR <sup>a</sup>
7'	88	3' UTR <sup>a</sup>				

Exon boundaries are shown in upper case and intron boundaries in lower case. Codon phase refers to the codon in the 5' end of the exon. <sup>a</sup>Exon 7' is repetitive element UTR.

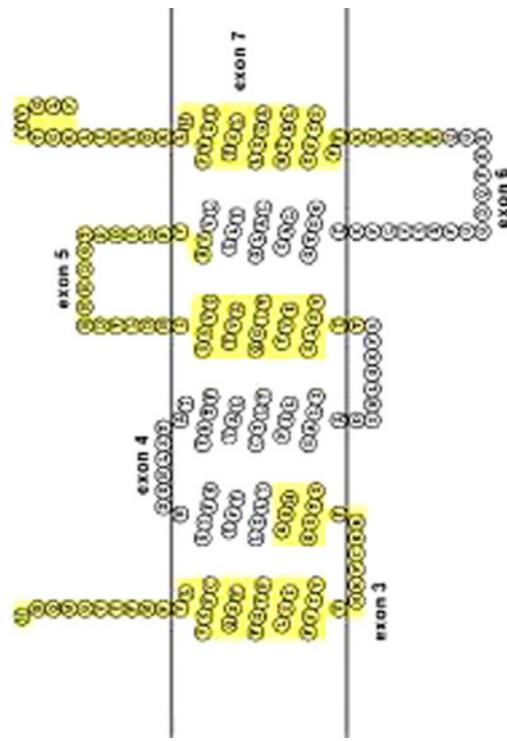


Figure 3. Membrane topology prediction for hPEPT1-RF. The prediction was carried out using the TOPPRED program. The exon boundaries are indicated by alternating color codes.

**Figure 4. Putative promoter region of the human PEPT1 gene. Sequence of nucleotides upstream of the translation start site (↑) (ATG) is shown. The numbering starts (+1) from the transcription start site (↑) to the negative values in the promoter region. The sites for the transcription factors in the promoter region are underlined and the corresponding indicated.**

```

-1964 tttcttaatc cgatcttatca ctgatggaca ttttagctgg tttccaagac attgctattg
-1904 tgaatagtgc ctcagtaaac atatgtgtgc acgtgtctt atagtagcat gattataat
-1844 ccttgggtg tatacccaat aatgggatcg ctgagtcaaa tggtagtttct agttctagat
                                         insulin
-1784 ccttgggaa ttgccacact gtctccaca atggtggaac taattccac tcccaccaac
-1724 agtgtaaaag tgccctatt cctccatatac ctctcagca ttgtgtgtt cctgactttt
-1654 taatgatcgc cattctaact ggcatgagat ggtatctgat tgggtttttt atttgcattt
-1594 ctctgatgac cagtatgtat gacatttt tcgtgtgtct gttggctgca taaatgtctt
                                         insulin
-1534 ctttgagaa gggctgttc atatccttg cccactttt gatgggtg tttgattttt
-1474 ttcttgtaaa tttgttaag ttctttgttag attctagata ttagtcctt gtcagatggg
                                         insulin
-1414 tagattgcaa acatttctc ccattctgta ggtgcctgt tcactctgat ggttagttct
-1354 ttgcgtgtgc agaagcttt tagtttaattt aggtcccatt tggcttattt ggctttgtt
-1294 gccattgctc ttgggtttt agacatgaag tccttgcaca tgcctatgtc ctgaatggta
-1234 ttgccttagt ttcttcttag gttttttagt attttaggtc ttacatcaa gtcttaatc
-1174 catttgaat taattttgt ataagggtta aggaaggat tcagtttcaat ctttctacat
-1114 atggctagcc agtttccca gcaccattta ttaaataggg aatccatttc ccatggctt
-1054 ttttgcgtat gtttgcataa gatcagatag ttgttagatgt gtgggtttat ttctgaggcc
-994 tctgttctgt tccattggc tatatatctg ttttggtaacc agtgcctatgc tgggttgc
-934 actgttagcct tggatgtat tttgaagtca ggttagcatga tgcctccagc tttgttctt
-874 ttgcttagga ttgtctggc tatggggct ctttttgggttccgtatgaa cttaaaagta
-814 gtttttcca attctgcgaa gaaaggcatt ggtagttga tgggcatacg attgaatcta
-754 taaattacct tgagcgtat gactatttt acgtatattga tccctcctat ccatgagcat
-694 agaatattct tccattggc tggatcctct taaagaaagg atgattctta aagaaaggaa
-634 atgagaatc ccccagaaat gcttcaaag gttgaatctc aaaataaaagc cacacacaca
                                         TATA box
-574 ctctcacaca cactctcaca cacacatcac acatacatcc tcacatgcaa actatata
                                         TATA box TATA box
-514 tattctcaca catgctcaca ctcaatcctc attcacacac acacactcac acacaatcc
-454 cactcacacc cacacacaca ttctctctca cacacacaca cacacacaca gctgaccact
-394 gtccccaccgt gtttgcctc cccacccaga agcgtgcgtt cctcccacag tggttccaa
-334 agtgcgttca tggatccagc agcatcaacc tccccttaga cttcttagaa atgcacattc
                                         GC-box GC-box
-274 tggggcccca gccccgacct cctgagtcag ctggccgggg ggtggggcc gggatccgc
                                         GC-box
-214 gtttagactg gctctccccc ggcctgccac ggcgtcccggt ggggcccac ccgcacctgg
-154 gcccgggctg gtgtccacgc ggcacggctc gggagcacgg acctctgcgc cccgcagcac
-94 cgccccccgg gtggagccgg cgccccggcc tcgcagagct ggggtgtac ctggggcaac
                                         GC-box
-34 ggggccggga ctggacgtca ggtcgagga gtagccacctg ccaggagcac gtccccgccgg
                                         ↑
                                         caggtcgacag gagccctggg agccggccatg
                                         ↑

```

All the exon-intron boundaries for *hPEPT1* conform to the consensus splice junction sequences (gt/ag) for eukaryotic genes (14). The 9 conserved nucleotides in the 5' donor side are (A/C)AG/gt(a/g)agt. These are conserved at 64%, 73%, 50%, 100%, 100%, 86%, 70%, 83%, and 77%, respectively, in *hPEPT1*. Similarly, we found positions of the 3'-acceptor site, (c/a)ag/(A/G), to be conserved at 73%, 100%, 100%, and 73%, respectively (Table 1). Splice sites are classified in phase 0 (13 sites), phase 1 (4 sites), and phase 2 (5 sites) (Table 1).

The *hPEPT1* gene structure shows several interesting features. The start sites of the transcripts for *hPEPT1* and pH-regulatory factor are located in different exons (Figure 1). Moreover, exon 1 located >20 kb upstream of exon 2 contains only the first 4 nucleotides of the *hPEPT1* coding region. Alternative splicing occurs in exon 3, and 118 bases in the 5' end of exon 3 are spliced out of the mRNA of *hPEPT1*. Another site for differential splicing is exon 7 of *hPEPT1-RF*. In this case, 41 bases in the 3' end of the exon are spliced out of *hPEPT1* mRNA (Figure 1).

Membrane topology predictions of *hPEPT1* and *hPEPT1-RF* proteins are shown in Figures 2 and 3. The transmembrane topology schematics were rendered using TOPO (S.J. Johns and R.C. Speth, Transmembrane protein display software, <http://www.sacs.ucsf.edu/TOPO/topo.html>, unpublished data). The figures show the peptide sequences that are encoded by each exon. In accordance with earlier information, *hPEPT1* is predicted to have 12 transmembrane segments (TMSs). Interestingly, comparison of membrane topology with gene structure shows possible functional modularity. Few if any exon-intron boundaries are found within the TMSs, and in most cases each exon encodes for a single TMS-loop unit (Figure 2). Topological predictions suggest that *hPEPT1-RF* has 5 TMSs with a cytoplasmic N-terminal and extracellular C-terminal.

The upstream region (2 kb) from the transcription start sites of *hPEPT1* is shown in Figure 4. TATA boxes were found about 520 bp upstream from the transcription start site in *hPEPT1*. The putative regulatory region also contains GC boxes, so several GC boxes are located within 300 bp from the transcription site in *hPEPT1*. Binding sites for transcription factors did not include any amino acid

responsive element. Some other transcription factor binding sites of the regulatory regions are illustrated in Figure 4.

## DISCUSSION

The genomic structure of *hPEPT1* and *hPEPT1-RF* presented here is based on a sequence in the HTGS database. The HTGS contains yet unordered pieces of genomic sequences. We used the August 11, 2000, version of the clone AL357553 in our analysis. It contained 11 contigs, but the true order of these pieces is still unknown, and the size of the gaps between them may change. Three introns of *hPEPT1* include such gaps (indicated by > signs in Table 1), while *hPEPT1-RF* exons are all located in one contig. Within the contigs the sequences are likely to be unaffected, and intron sizes are reliable (Table 1). Note also that the order of the exons in the clone matches perfectly the nucleotide sequence of cDNA. Where possible, these predictions have subsequently been verified and the intronic sizes adjusted where needed, on the basis of additional genomic clones deposited in the HTGS database.

Human *PEPT1* is encoded by 23 exons, and the entire gene contains 24 exons. Likewise, mouse *PEPT1* is encoded by 23 exons (13). Comparison of mouse and human genes shows that the sizes of the exons and their relative locations are similar. Identity of mouse and human cDNA for *PEPT1s* is 83% (13). A high degree of similarity in both gene clustering and coding sequence confirm that human and mouse *PEPT1* genes are orthologues. In this study the comparison of membrane topological prediction and genomic structure indicates that human *PEPT1* gene is modular with each TMS-loop unit encoded by a different exon (Figures 2 and 3). This is in accordance with earlier analysis of peptide transporters that suggested modular structure of transporter genes may have evolved by exon shuffling and rearrangements of functional modules (15).

The *hPEPT1* gene also encodes the splice variant *hPEPT1-RF*. *PEPT1-RF* and *PEPT1* share 5 identical TMSs, while the extramembranous terminals differ (Figures 1-3). *PEPT1-RF* is not capable of transporting substrates across the membrane, but it is thought to sense pH changes and modulate the response of *PEPT1* to these changes (12). Fei et al (16) have shown by using chimeric *PEPT1-PEPT2* proteins that the TMSs 7-9 are important for substrate recognition by *hPEPT1*. *PEPT1-RF* does

not have these TMSs and does not transport substrates. However, the mechanisms of proton and substrate transfer and the interplay between PEPT1 and PEPT1-RF are still elusive.

The putative regulatory region of *hPEPT1* (Figure 4) revealed some similarities with the mouse *PEPT1* gene (13). TATA boxes are located in unusual locations (511 bp and 517 bp upstream from the transcription start site), while GC boxes are located near the start site (at -29 bp and several others within 300 bp). The location of TATA boxes so far upstream from the transcription start site is not optimal. Therefore, this kind of structure suggests that the GC box is a more important promoter in the regulation of *hPEPT1* than is the TATA box. Note also that there may be more than one transcription start site for a gene, as shown previously (17). Unlike in the mouse genome, amino acid responsive element was not found within 1983 bp from the transcription start site in *PEPT1*. Human PEPT1 expression is known to be upregulated by its substrates, dipeptides, as shown by Walker et al (18), but the mechanism of this upregulation remains unclear. Insulin regulates the activity of PEPT1 in Caco-2 cells (19). Insulin regulation was mediated by transporter translocation to the basolateral side of the cells upon release of hPEPT1 from the translated intracellular pool to the plasma membrane. Changes in hPEPT1 mRNA were not seen in that study. However, the putative insulin responsive element is located upstream from the transcription start site (Figure 4), suggesting that insulin might be involved in the regulation of *hPEPT1* transcriptional activity.

The genomic organization of *hPEPT1* and *hPEPT1-RF* indicates that they are splice variants of the same gene (Figure 1). Expression of hPEPT1-RF has not been studied in detail. Nevertheless, the splice variants may be expressed in different proportions depending on, for example, the stage of differentiation, hormonal regulation signals, and cell type. Human PEPT1 is expressed in several tissues (intestine, kidney, brain, liver) where the pH environment is quite different. Also, an intracellular pool of hPEPT1 may be associated with peptide trafficking in lysosomes and endosomes that have different pH depending on the maturity of the vesicle (11).

Finally, the genomic organization of *hPEPT1* paves the way for studies of the relationships between *PEPT1* genotype and pharmacokinetics.

## ACKNOWLEDGEMENTS

This study was partly supported by the Academy of Finland (AU) and a grant from the National Institutes of Health, Bethesda, MD, General Medical Sciences (GM99004).

## REFERENCES

1. Sadée W, Druebbisch V, Amidon GL. Biology of membrane transport proteins. *Pharm Res*. 1995;12:1823-1837.
2. Fei YJ, Kanai Y, Nussberger S, et al. Expression cloning of a mammalian proton-coupled oligopeptide transporter. *Nature*. 1994;368:563-566.
3. Liang R, Fei YJ, Prasad PD, et al. Human intestinal H<sup>+</sup>/peptide cotransporter: cloning, functional expression, and chromosomal localization. *J Biol Chem*. 1995;270:6456-6463.
4. Liu W, Liang R, Ramamoorthy S, et al. Molecular cloning of PEPT2, a new member of the H<sup>+</sup>/peptide cotransporter family, from human kidney. *Biochem Biophys Acta* 1995;1235:461-466.
5. Yamashita T, Shimada S, Guo W, et al. Cloning and functional expression of a brain peptide/histidine transporter. *J Biol Chem*. 1997;272:10205-10211.
6. Botka C, Wittig TW, Graul RC, et al. Human proton/oligopeptide transporter (POT) genes: identification of putative human genes using bioinformatics. *AAPS PharmSci*. 2000;2: (2), article 16.
7. Covitz KMY, Amidon GL, Sadée W. Membrane topology of the human dipeptide transporter, hPEPT1, determined by epitope insertions. *Biochem*. 1998;37:15214-15221.
8. Shen H, Smith DE, Yang T, Huang YG, Schnerman JB, Brosius FC. Localization of PEPT1 and PEPT2 proton coupled oligopeptide transporter mRNA and protein in rat kidney. *Am J Physiol*. 1999;276:F658-F665.
9. Ganapathy V, Leibach FH. **Transport Across Intestinal Epithelium** In: Yamada T, ed. *Textbook of Gastroenterology*. Philadelphia, PA: Lippincott Williams and Wilkins; 1999:456-467.
10. Terada T, Sawada K, Saito H, Hashimoto Y, Inui KI. Functional characteristics of basolateral peptide transporter in the human intestinal cell line Caco-2. *Am J Physiol*. 1999;276:G1435-G1441.
11. Gonzales DE, Covitz KMY, Sadée W, Mrsny R. An oligopeptide transporter is expressed at high levels in the pancreatic carcinoma cell lines AsPc-1 and Capan-2. *Cancer Res*. 1998; 519-525.
12. Saito H, Motohashi H, Mukai M, Inui K. Cloning and characterization of a pH-sensing regulatory factor that modulates transport activity of the human H<sup>+</sup>/peptide cotransporter, PEPT1. *Biochem Biophys Res Commun*. 1997;237:577-582.
13. Fei YJ, Sugawara M, Liu JC, et al. cDNA structure, genomic organization, and promoter analysis of the mouse intestinal peptide transporter PEPT1. *Biochem Biophys Acta*. 2000;1492:145-154.
14. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl Acids Res*. 1987;15:7155-7174.
15. Graul RC, Sadée W. Sequence alignments of the H<sup>+</sup>-dependent oligopeptide transporter family PTR: inferences on structure and function of the intestinal PEPT1 transporter. *Pharm Res*. 1997;14:388-400.
16. Fei YJ, Liu JC, Fujita T, Liang R, Ganapathy V, Leibach FH. Identification of a potential substrate binding domain in the mammalian peptide transporters PEPT1 and PEPT2 using PEPT1-PEPT2 and

PEPT2-PEPT1 chimeras. Biochem Biophys Res Commun. 1998;246:39-44.

17. Pavé-Preux M, Aggerbeck M, Veyssier C, Bousquet-Lamercier B, Hanoune J, Barouki R. Hormonal discrimination among transcription start sites of aspartate amino transferase. *J Biol Chem.* 1990;265:4444-4448.

18. Walker D, Thwaites DT, Simmons NL, Gilbert HJ, Hirst BH. Substrate upregulation of the human small intestinal peptide transporter, hPept1. *J Physiol.* 1998;507:697-706.

19. Thamotharan M, Bawani SZ, Zhou X, Adibi SA. Hormonal regulation of oligopeptide transporter PepT1 in a human intestinal cell line. *Am J Physiol.* 1999;276:C821-C826.

## APPENDIX

```
>gene exon1  
CCACCTGCCAGGAGCACGTCCGCCGGCAGTCGCAGGAGCCCTGGAGCCGCCATGG  
  
>gene exon2  
GAATGTCCAATCACAC  
  
>gene exon3  
CAGTCACCGGTAAGTTACTGAATCTGCGTTGGCTGCCTTCATGCCACTGGTGCCTT  
GTGGCTATGGACCAGTGGGGGTGTATTCATCCTGTGCTTCTTGTCAAGT  
TTCTTGTTATCCCCTGAGCATCTTCATCGTGGTCAATGAGTTGCGAAAGATT  
TCCTACTATGGAATGCGAG  
  
>gene exon 4  
CAATCCTGATTCTGTACTTCACAAATTTCATCAGCTGGATGATAACCTGTCCACCGCA  
TCTACCATACTGTTGTGGCTGTGCTACCTGACGCCAATTCTGGAGCTTATGCCG  
ACTCGTGGCTGGAAAGTTCAA  
  
>gene exon 5  
GACCATTGTGTCGCTCTCCATTGTCTACACAATTGGACAAGCAGTCACCTCAGTAAGCTC  
CATTAATGACCTCACAGACCACAACCATGATGGCACCCCCGACAGCCTCCTGTGACGT  
  
>gene exon 6  
GGTGTCTGCCCTTGATCGGCCTGCCCTGATAGCTCTGGGACTGGAGGAATCAAACCTG  
TGTGTCTGCCCTGGAGATCAGTTGAAGAGGGCAG  
  
>gene exon 7  
GAGAAACAAAGAACAGATTTTTCCATTTTACTTGGCTATTAATGCTGGAAGTTG  
CTTCCACAATCATCACACCCATGCTCAGAGGTAAAGAGATACTGAAAGGAACCTCTGTT  
GGTCTTTTAGTAAGGCCAAGTCCGTCTTCTCGTTGTCTGTTCTCCCTCTG  
CCCTGTTCTAGCTGTCAGCTACATGCTGGATGCACGTTGAATCTGGAGTAAGTCT  
TCCTCCCTGTGAGTTCACCTCTTTGTCTTAAATCCTTATCCCAGTGA  
ACTGGAGAGTCTAACCCATCCTAGCATGGCATCCATCTCTAAACTGTTATTCA  
AAAGTCATGCTGTTGAAGAGGTAAAGAGGCCAAGAGGGTGCCTTGAAGCCAAGCTGGT  
GCCGTAGCTGCCATTAGCCTCAAGCTTGTCTCCCCAGTGATATTCTGTGACTTAG  
AAGAGTTACTGGAGTCAAAATTGATTGATTTTGTGTTGTTGTTGTTAGAAG  
GAGTCTTGTCTGTCAACCCAGCTGGAGGCAGTGGCGTGTCTGGCTACTGCAACCTC  
CACCTCCTGGTTCAAGCAATTCTCCTGCCTCAGCCTCTGAGTAGCTGGACAACAGGC  
GCCTGCCACCATGCCCGATAATATTTTGTTAGTAGAGATGGGATTCACCAT  
GTTAGCCAGGATGGCTCGATCTCCTGCCCTGTGATCTGCCTGCCTGCCCTCCAAAG  
TGCTGGGATTACAGGTGTGAGCACCACGCCCTGGCTAAGGCAAAGTTTTCTTGCTT  
TCCCATCATCTCAAGTATTCCAAAAGAACTGTCAGATAAAGAGAATGGACAGTGTCA  
GTAACATCAGAGGTTGTGCTGCTAATGCTCATAGAGAGTTGTGAATTCCCATGT  
GGAAGGGAATATGTGGTTACTGGTACAAACTAACAGACCATCTGTCCACATTGA  
ATCAGATTAGATTATGAAATAAGATTCTCAGGAGA  
  
>gene exon 8  
TTCAACAATGTGGAATTCACAGTAAACAAGCTTGTACCCACTGGCTTGGGTTCTG  
CTGCTCTCATGGCTGTAGCCCTGA  
  
>gene exon 9  
TTGCCATAAAAATAGATTTAGGCATGGAGTAAGGCATTCCAAGAGGGAGCACTGG  
GTAAAGTGGCAAGTGCATCGGT  
  
>gene exon 10  
TTGCCATCAAAATAGATTTAGGCATGGAGTAAGGCATTCCAAGAGGGAGCACTGG  
CTGGACTGGCTAAAGAGAAATACGAT  
  
>gene exon 11  
GAGCGGCTCATCTCCAAATTAAGATGGTACGAGGGTGTGTTCTGTATATTCCACTC
```

CCAATGTTCTGGGCCTTGTGACCAGCAG

>gene exon 12

GGCTCCAGGTGGACACTGCAGGCAACAACATATGTCGGGAAAATC

>gene exon 13

GGAGCTCTGAAATTCAGCCGATCAGATGCAG

>gene exon 14

ACCGTGAAACGCCATCTGATCGTATGGTCCCCTTCGATGCTGTGCTGTACCC  
CTCATTGCAAAATGTGGCTCAATTCA

>gene exon 15

CTCCTTGAAGAAGATGGCAGTGGCATGGCCTGGCCTCCATGCCCTTGTGGTGGCTGC  
CATCGTGCAGGTGGAAATCGAT

>gene exon 16

AAAACCTTCCAGTCTCCCCAAAGGAAACGAAGTCAAATTAAAGTTGAATATAGGA  
AACAAATACCATGAATATATCTCTCCTGGAGAGATGGTACACTGGCCAATGTCTCAA

>gene exon 17

ACAAATGCATTTATGACTTTGATGTAAACAAACTGACAAGGATAAACATTTCTCCT  
GGATCACCAAGTCACTGCTGTAAGTGACGACTTCAAGCAGGGCAACGCCACACGCTTCTA  
GTGTGGGCCCAATCACTACCAGGTG

>gene exon 18

GTAAAGGATGGCTTAACCAGAAGCCAGAAAAAGGGAAAATGGAATCAG

>gene exon 7'

ATTGCTTGAACCTGGTAGACGGGGTTGCAGTGAGCTGAGATCGTCTGCTGCACTCCAG  
CCTAGCCCACAGAGCAAGACTCTGTCTC

>gene exon 19

ATTGTAAATACTTTAACGAGCTCATCACCATCACATGAGTGGAAAGTTATGCAA  
CATCAGCAGCTACAATGCCAGCACATACCAGTTTCTGGCAT

>gene exon 20

AAAAGGCTTCACAATAAGCTAACAGAGATTCCGCCAACATGTCACCTAACATAC  
TTTCTACCTTGAATTGGTAGTGCTTACCTATAGTCAAAGGAAG

>gene exon 21

ATGACAGCTGCCCTGAAGTGAAGGTGTTGAAGATATTCAGCCAACACAGTTAACATG  
GCTCTGCAAATCCCGAGTATTCTCACCTGTGGCGAAGTGGCTCTGTACG  
GGATTGGAATTCTCATATTCTCAG

>gene exon 22

GCTCCTTCAACATGAAGTCGGTCTTCAGGCAGGATGGCTGCTGACCGTGGCTGTTGGC  
AACATCATGTGCTCATCGTGGCAGGGCAGGCCAGTCAGCAAACAG

>gene exon 23

TGGGCCGAGTACATTCTATTTGCCGCGTTGCTTCTGGTCGTCTGTGAATTTGCCATC  
ATGGCTCGTTCTACTTACATCAACCCAGCGGAGATCGAAGCTCAATTGATGAGGAT  
GAAAAGAAAAACAGACTGGAAAAGAGTAACCCATATTCTAGTCAGGGCCAATTACAG  
AAACAGATGTGAAGGTCAAGGAGCAAGTGGAGGATGGACTGGGCCCGCAGATGCCCTGAC  
CTCTGCCCTCAGGTAGCAGGACACTCCATTGGATGGCCCTGATGAGGAAGACTTCAGAA  
TTGGGAACTAACCATGAATGCTATTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT  
TTTTTTGAGACAGAGTTGCTCTGTTGCTCAGGCTGGAGTGCAATGGCACGATCTCA  
GCTCACTGCAACCTCCGCCCTCCAGGTTCAAGTAATTCTCTGCCAGCCTCCGAGTG  
GCTGGGATTAGCGGCATGCACCACGCCAGCTATTGTTAGTAGAGATGG  
GGTTTCACCATGTTAACAGGATGGCTCGATCTTGACCTGGTATCTGACCGTACCG  
GCCTGCCAAAGTGTGGATTACAGGCTTGAGCTACCGCGCCCGTGAACGCTATT

TCTAAGCAGCCAGCAGTGAATCTAAAACCTGGAAAGAAGTCTTCTGTTGAAAGGCTTAT  
TTAACGCCACACGTACACACACTGTCTTAGAGTACTGTGAGCCCACCCCCACATTGGTCATC  
TTCCCTATCACACAAATGATGTTATTTGACTAGCTTAATTGAAATGTAACAAAGT  
TTCCTATTCCATACTGTTCAATTACTCTTACGAAAACATTCTAAAGGAGGCAGG  
AGCCAAGGCCAAAAGTGAACGTACAGGTTAAAATGGCTGTGATAAGGGCAGCTGGTAT  
TAACTGATAACTTACCTTGGTTTTGTTATTTGTTCTAGTCCCTACCTGTGTT  
TAAATTATGGATAACTCGAAAGACAGGCTCAGGTGAAGGCCGAGTAATGATTTTTGAA  
AGTTTCAATGGTGTGAAATAAATTCTGTTCTTA

---

>hpept1 exon 1  
CCACCTGCCAGGAGCACGTCCCAGGCAGTCGAGGAGCCCTGGAGCCGCCATGG

>hpept1 exon 2  
GAATGTCAAATCACAC

>hpept1 exon 3  
AGTTTCTTGTTATCCCCCTGAGCATCTTCTTCATCGTGGTCAATGAGTTGCGAAAGA  
TTTCCTACTATGGAATGCGAG

>hpept1 exon 4  
CAATCCTGATTCTGTACTTCACAAATTTCATCAGCTGGATGATAACCTGTCCACCGCCA  
TCTACCATACTGTTGTGGCTCTGTGCTACCTGACGCCAATTCTCGGAGCTTATGCCCG  
ACTCGTGGCTGGAAAGTTCAA

>hpept1 exon 5  
GACCATTTGTGCGCTCTCCATTGTCTACACAATTGGACAAGCAGTCACCTCAGTAAGCTC  
CATTAATGACCTCACAGACCACAACCATGATGGCACCCCCGACAGCCTCCTGTGCACGT

>hpept1 exon 6  
GGTGTCTGCCCTGATCGGCCTGGCCCTGATAGCTCTGGGACTGGAGGAATCAAACCTG  
TGTGTCTGCGTTGGAGATCAGTTGAAGAGGGCCAG

>hpept1 exon 7  
GAGAAACAAAGAAACAGATTTTTCCATTTTACTTGGCTATTAATGCTGGAAGTTG  
CTTTCCACAATCATCACACCCATGCTCAGAG

>hpept1 exon 8  
TTCAACAATGTGGAATTACAGTAAACAAGCTTACCCACTGGCCTTGGGTTCTG  
CTGCTCTCATGGCTGTAGCCCTGA

>hpept1 exon 9  
TTGTGTTTGTCCCTGGCAGTGGATGTACAAGAAGTTCAAGCCACAGGGAACATCATGG  
GTAAAGTGGCCAAGTGCATCGGT

>hpept1 exon 10  
TTTGCCATAAAAATAGATTAGGCATCGGAGTAAGGCATTCCAAGAGGGAGCACTGG  
CTGGACTGGCTAAAGAGAAATACGAT

>hpept1 exon 11  
GAGCGGCTCATCTCCAAATTAAAGATGGTACGAGGGTGTGCTGTATATTCCACTC  
CCAATGTTCTGGCCTTGTGACAGCAG

>hpept1 exon 12  
GGCTCCAGGTGGACACTGCAGGCAACAACTATGTCGGAAAATC

>hpept1 exon 13  
GGAGCTCTGAAATTCAAGCCGATCAGATGCAG

>hpept1 exon 14  
ACCGTGAACGCCATCTGATCGTACGGTCCGATCTCGATGCTGTGCTGTACCC  
CTCATTGCAAAATGTGGCTTCAATTTCAC

>hpept1 exon 15  
CTCCTTGAAGAAGATGGCAGTTGCATGGCCTGGCCTCATGGCCTTGTGGTGGCTGC  
CATCGTGCAGGTGGAAAATCGAT

>hpept1 exon 16  
AAAACCTCTCCAGTCTCCCCAAAGGAAACGAAGTCAAATTAAAGTTGAATATAGGA  
AACAAATACCATGAATATCTCTTCTGGAGAGATGGTACACTTGGCCAATGTCTCAA

>hpept1 exon 17  
ACAAATGCATTATGACTTTGATGTAAACAAACTGACAAGGATAAACATTCCTCCT  
GGATCACCACTGACTGCTGTAACTGACGACTTCAAGCAGGCCACACGCTTCTA  
GTGTGGGCCCAATCACTACCAGGTG

>hpept1 exon 18  
GTAAAGGATGGTCTTAACCAGAAGCCAGAAAAAGGGAAAATGGAATCAG

>hpept1 exon 19  
ATTGTAAATACTTTAACGAGCTCATCACCATCACATGAGTGGAAAGTTATGCAA  
CATCAGCAGCTACAATGCCAGCACATACCAGTTTCCTCTGGCAT

>hpept1 exon 20  
AAAAGGCTTCACAATAAGCTCACAGAGATTCCGCCAACATGTCAACCTAATTCAATAC  
TTTCTACCTTGAATTGGTAGTGCTTACCTATAGTCCAAGGAAG

>hpept1 exon 21  
AATGACAGCTGCCCTGAAGTGAAGGTGTTGAAGATATTCAGCCAACACAGTTAACATG  
GCTCTGCAAATCCCGAGTATTCTCACCTGTGGCGAAGTGGCTCTGTCACTG  
GGATTGGAATTCTCATATTCTCAG

>hpept1 exon 22  
GCTCCTTCAACATGAAGTCGGTCTTCAGGCAGGATGGCTGCTGACCGTGGCTGTTGGC  
AACATCATTGTGCTCATCGTGGCAGGGCAGGCCAGTCAGCAAACAG

>hpept1 exon 23  
TGGGCCGAGTACATTCTATTGCCGCGTTGCTCTGGTCGTCTGTGAATTTTGCCATC  
ATGGCTCGTTCTATACTTACATCAACCCAGCGGAGATCGAAGCTCAATTGATGAGGAT  
GAAAAGAAAAACAGACTGGAAAAGAGTAACCCATATTTCATGTCAGGGCCAATTACAG  
AAACAGATGTGAAGGTCAAGGAGCAAGTGGAGGATGGACTGGCCCGCAGATGCCCTGAC  
CTCTGCCCTCAGGTAGCAGGACACTCCATTGGATGGCCCTGATGAGGAAGACTTCAGAA  
TTGGGAACTAAACCATGAATGCTATTTCCTTTCTTTCTTTCTTTTTTTTT  
TTTTTTGAGACAGAGTTGCTTGTCCAGGCTGGAGTGAATGGCACGATCTCA  
GCTCACTGCAACCTCCGCCCTCCAGGTTCAAGTAATTCTCCTGCCCTCAGCCTCCGAGTG  
GCTGGGATTAGCGGCATGCACCACACGCCAGCTATTGTATTAGAGATGG  
GGTTTCACCATGTTAACAGGATGGCTCGATCTCTGACCTGGTATCTGCCACCTCG  
GCCTGCCAAAGTGTGGATTACAGGCTTGAGCTACCGGCCGGCGTGAACGCTATT  
TCTAAGCAGCCAGCAGTGAATCTAAACTCTGGAAGAAGTCTCTGTTGAAAGGCTTAT  
TTAAGCCACACGTACACACACTGTCTTAGAGTACTGTGAGGCCACCCACATTGGTCATC  
TTCCCTATCACACAAATGATGTTATTGGACTAGCTTAATTGAAATGTAACAAAGT  
TTCCTATTCCATACTGTTCAATTCTAAACTCTTACGAAAATCTAAAGGAGGCAGG  
AGCCAAGGCCAAAAGTGAACGTACAGGTTAAATGGCTGTGATAAGGCCAGCTGGTAT  
TAACTGATAACTTACCTTGGTTTTGTTATTGTTCTAGTCCCTACCTGTGTT  
TAAATTATGGATAACTCGAAAGACAGGCTCAGGTGAAGGCCAGTAATGATTTTTG  
AGTTCAATGGTGTGAAATAAATTCTGTTCTTA

>hpept1-rf exon 3

CAGTCACCGGTAAGTTACTGAATCTGCCTGGCTGCCCTTCACTGCCACTGGTGCTCCTT  
GTGGCTATGGACCAGTGGGGGGTGTATTCATCCTGTGCTTCTTGTCAAGAGT  
TTCTTGGTTATCCCCTGAGCATCTTCATCGTGGTCAATGAGTTTGCAGAAAGATT  
TCCTACTATGGAATGCGAG

>hpept1-rf exon 4

CAATCCTGATTCTGTACTTCACAAATTCATCAGCTGGATGATAACCTGTCCACCGCCA  
TCTACCATACTGTTGTGGCTCTGTGCTACCTGACGCCAATTCTCGGAGCTTATGCCG  
ACTCGTGGCTGGAAAGTTCAA

>hpept1-rf exon 5

GACCATTGTGTCGCTCTCATTGTCTACACAATTGGACAAGCAGTCACCTCAGTAAGCTC  
CATTAATGACCTCACAGACCACAACCATGATGGCACCCCGACAGCCTCCTGTGCACGT

>hpept1-rf exon 6

GGTGTGCTGCCCTTGATCGGCTGCCCTGATAGCTCTCGGACTGGAGGAATCAAACCTG  
TGTGTCTGCGTTGGAGATCAGTTGAAGAGGGCCAG

>hpept1-rf exon 7

GAGAAACAAAGAACAGATTTTCCATCTTTACTTGGCTATTAATGCTGGAAGTTG  
CTTTCCACAAATCATCACACCCATGCTCAGAGGTAAAGAGATACTGAAAGGAACCTCTGTT  
GGTCTTTTTAGTAAGGCCAAGTCCGTCTTCTCGTTGTCTGTTCTCCCTCTG  
CCCTGTTCTAGCTGTCAGCTACATGCTGGATGCACGTGGAATCTGGAGTAAGTGC  
TCCTTCCCTGTGAGTTCACCTCTTTGTCTTAAATCCTTATCCCAGTGACC  
ACTGGAGAGTCTAACCCATCCTAGCATGGCATCCATCTCCTAAACTGTTATTGAC  
AAAGTCATGCTCGTGTAAAGAGGTAAAGAGGCCAAGAGAGGTGCCCTGGAAGCCAAGCTGGT  
GCCGTAGCTGCCATTAGCCTCAAGCTTGCTCCCCAGTGATATTCTGTGACTTAG  
AAGAGTTACTGGAGTCAAAATTGATTGATTTTGTGTTGTTGTTGTTGTTAGAAG  
GAGTCCTGCTCTGTCACCCAGCCTGGAGGCAGTGGCGTGTACTGGCTCACTGCAACCTC  
CACCTCCTGGTTCAAGCAATTCTCCTGCCTCAGCCTCTGAGTAGCTGGACAACAGGC  
GCCTGCCACCATGCCCGATAATATTTTGTTAGTAGAGATGGGATTCACCAT  
GTTAGCCAGGATGGCTCGATCTCCTGCCCTCGTGTACTGCCCTGCCCTCCAAAG  
TGCTGGGATTACAGGTGTGAGCACCACGCCCTGGCTAAGGCAAAGTTTTCTTGCTT  
TCCCATCATCTCAAGTATTCAAAAGAACTGCTCAGATAAAGAGATGGACAGTGTCA  
GTAACATATCAGAGGTTGTGCTGCTAATGCTCATAGAGAGTTGTGAATTCCCATGT  
GGAAGGAAATATGTGGTGTACTGGTACAAACTAACAGACCACCTGTCCACATTGA  
ATCAGATTAGATTATGAAATAAAGATTCAAGGAGA

>hpept1-rf exon 7' 3' UTR

ATTGCTTGAACCTGGTAGACGGGGTTGCAGTGAGCTGAGATCGTGTGCTGCACTCCAG○  
CCTAGCCCACAGAGCAAGACTCTGTCTC